Definition of Stemma and Archetype

Roelli, Philipp

DOI: https://doi.org/10.1515/9783110684384-005

variant is to be understood in relation to constructing the stemma; the second categorisation addresses the different sorts of variation and their potential causes, relating them to Quintilian's four main categories of error (addition, omission, substitution, and transposition).

Although the text-genealogical principles behind stemmatology seem straightforward, even obvious, when they are first encountered, a philologist confronted with real historical texts will soon encounter complications. Foremost among these is the phenomenon of the so-called contaminated witness, which is to say, a text manuscript that was copied with reference to more than one exemplar. Tuomas Heikkilä treats this subject in section 4.4, where he demonstrates how contamination can lead to erroneous stemmata, describes the different modes in which a text might have been copied from multiple sources, and provides some guidelines for how an editor might deal with the situation, gaining insight into the transmission history of the text even if a complete and definitive stemma cannot be drawn.

The transmission history that is represented by a stemma is the subject of section 4.5, the last in this chapter, by Caroline Macé. Here, the reader is treated to a demonstration of the need to study the history of a text not only on the basis of its variant readings, but also in light of the paratextual and contextual knowledge that we have about the documents that carry the text. Macé presents three case studies, each of which shows in a different way the inadequacy of restricting oneself either to historical analysis or to stemmatic analysis. The third case study also discusses issues that arise when the text under examination exists only (or primarily) in translation, which may forestall the use of automated collation software but requires the editor nevertheless to find a way to carry out meaningful comparison of texts in different languages.

In sum, this chapter contains a great deal of information, sometimes presented in an unavoidably dense manner, about the form, function, and significance of what in many cases appears to be a simple diagram. Only with a full understanding of the concepts and complications covered in this chapter, however, can the reader avoid the pitfalls of a naive use of the computational methods that follow in chapter 5.

4.1 Definition of stemma and archetype

Philipp Roelli

This section considers the two key concepts for the genealogical reconstruction of texts, already mentioned in passing in previous chapters, in more depth: stemma and archetype. Their historical context, application, types, and definitions will be examined. The next section will then consider the stemma as a computational model.

4.1.1 Context

"Stemma" and "archetype" are probably the two most important terms in traditional genealogical textual philology. After some preliminary remarks, more formal definitions will be proposed. As a first approximation, one may imagine the stemma as the genealogical tree of all known, extant witnesses of a text and the archetype as their most recent common ancestor, usually lost. In practical terms, the archetype is the uppermost point in a stemma, on which all extant branches converge (on the relation between archetype and original, see 4.1.4), or, seen from the other end, the point beyond which *recensio* of the extant tradition of a text cannot reach (see Trovato 2005, 12). Originally, the main point of devising a stemma for a textual tradition was to reduce the amount of possible choice between variants for its editor: the stemma can in many cases show that a reading was innovated and could not have stood in the archetype (see 2.3.2 for Gaston Paris on this topic). Today, stemmata are also used in many other contexts when studying the transmission of a text. Often, editors who wish to edit a text as closely as possible to the original try to reconstruct the archetype's text as far as possible (see 2.2). But it is crucial to be aware that the archetype is usually not identical with the author's original text – in fact, many centuries may lie between these two texts. The archetype may be any witness that acquired this special and important function in the transmission of its text by historical chance; indeed, it may be a witness full of mistakes and deficiencies of all kinds. Faced with a faulty archetype of this kind, the editor will usually try to improve the archetypal text using external data or conjecture (see 6.2.3 on the delicate task of *emendatio*). If there are more than a very few witnesses, the reconstruction of a stemma is usually not a trivial task and is often disputed among editors of the same text. New insights into the text's transmission and significant changes in the stemma can necessitate an entirely new critical edition. As a rule of thumb, it may be said that, the more witnesses there are, the more difficult it becomes to figure out all relationships between them and to draw an adequate stemma; this problem is aggravated by the fact that the probability of contamination (see 4.4) increases as the number of witnesses does. In some cases, for instance if there are a great number of witnesses – there may be hundreds, occasionally even thousands – it may not be feasible to construct a stemma at all (see 7.1 for examples).

4.1.2 History of the terms

The Latin expression *stemma codicum*, or in short just *s t e m m a* (plural *stemmata*), literally means "genealogical tree of the manuscripts". The word "stemma" ultimately derives from the Greek word στέμμα (pl. στέμματα), "wreath, garland",

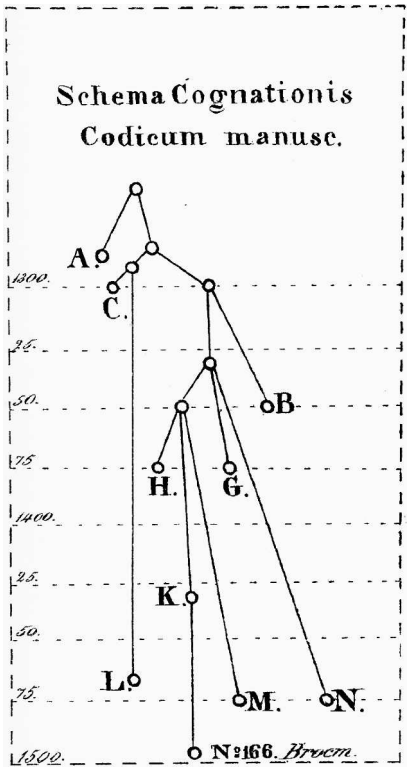

Fig. 4.1-1: Schlyter's *schema cognationis* for the *Västgötalagen* (Schlyter and Collin 1827, appendix), which may be the first printed stemma.

which is derived from the verb στέφω, "put/hang around". It is already used figurativity in Latin Antiquity to mean "genealogical tree" (e.g. in Suetonius *De vita caesarum*, Claudius 2). But what we today call a stemma in textual criticism is a recent acquisition: the idea was apparently proposed for the first time in the eighteenth century by Bengel in the context of a hypothetical genealogical tree of witnesses of the New Testament, although he did not use the name (he called it a "tabula quaedam quasi genealogica"; Bengel 1763, 20 [a certain, so to speak, genealogical table]). Apparently, it was only in the nineteenth century that such *tabulae* were first printed in editions; the first scholar to print a stemma may have been Carl Johan Schlyter in 1827 (Schlyter and Collin 1827, appendix; he called it a *schema cognationis codicum manuscriptorum* [diagram of relationship of the manuscripts], see fig. 4.1-1), whereas Carl Gottlob Zumpt (1831, xxxviii) may have been the first person to use the designation *stemma codicum manuscriptorum* in 1831 (see Timpanaro 1961, 61). Nevertheless, he still relegated the actual stemma to a footnote. The term becomes the accepted technical term in the wake of Paul Maas's *Textkritik* (1927). Some more details about the early history of scholarly stemmata can be found below in section 6.1.2. When prints, not manuscripts, are the witnesses to be discussed, the full Latin term *stemma editionum* is sometimes used (see examples in 7.8).

The word "archetype" is derived from the classical Greek compound ἀρχέτυπον, "archetype, pattern, model, exemplar", which was often opposed to ἀπόγραφον, "copy" (see 1.1.5). The compound itself consists of ἀρχή, "beginning", and τύπος, "the effect of a blow or of pressure" and thus "impression, seal, engraving, etc.". Renaissance scholarship (written in Latin) tended to use the word *archetypus* in the classical Latin sense as "autograph" (Irigoin 1977); this may cause confusion, as the modern scholarly meaning is rather different. In reality, the situation is even more complicated; Rizzo (1973, 308–318) differentiates at least four different Renaissance meanings of "archetype" and studies the history of the term further.

4.1.3 The *stemma codicum*

The basic, practical method of arriving at a stemma (*constitutio textus*), including a fictitious and a real example, has already been presented above (see 2.2.4–6). The general idea described there can be formalised into a definition such as the one we propose here: a *stemma (codicum)* is an oriented tree-like graph representing a hypothesis about genealogical relationships between witnesses of a text.

This definition uses the terms "tree-like", "graph", "'witness", and "text", some of which come from a traditional philological background, others from a mathematical one. The philological concept of a witness was discussed in section 2.2; while we refrain from attempting to define here the elusive term "text", some examples of the sometimes fluid boundaries of "texts" are provided in section 3.2. On the other hand, the terms "tree" and "graph" are mathematical ones. A purely computational approach to the concept of the stemma is presented in the next section (4.2). There, the Greg tree as a mathematically defined version of the traditional stemma without contamination (4.2.3.3) is introduced. Section 5.2 explains what "tree" and what the more general term "graph" mean in mathematics; the related terms "DAG" and "polytree" are also introduced there. The term "tree-like" is intentionally fuzzy: it is intended to hint at the fact that the graph can be turned into a tree by removing some edges, the ones accounting for contamination (see 4.4). The defining characteristic of a tree is that any two nodes are connected by exactly one path; this holds only in the traditional, uncontaminated situation in which all witnesses are copied from only one ancestor each. Put another way, if only the main line of descent for each witness is used, the stemma will become a tree. If all transmission of information between witnesses is included, and if there was contamination, the diagram will no longer be a tree. For example, if one witness, say *C*, is a copy of another witness, say *A*, then the graph depicting the two is oriented; in this case, the direction is from *A* to *C*. But if witness *C* was copied partly from *A* and partly from another witness, say *B*, then there will be two (or more) paths from the archetype to witness *C* (one through *A* and the other through *B*), and the stemma will no longer be a tree. In order to turn it back into a tree, either *A* or *B* would have to be regarded as the main line of descent, and the other path would have to be suppressed.

In order to arrive at the stemma most accurately depicting the historical transmission, a philologist will use all available information about the witnesses, including but not limited to the text-state they carry. There will be additional information about the witnesses as objects, such as palaeographical estimates of their age, the identification of different hands writing the text, or information gleaned from the page layout. And, with luck, additional information may also be contained in colophons (see 1.2.2) – scribes may explicitly state what their sources were, when and where they wrote, for instance. For editors, the stemma is crucial in that it helps them reconstruct the archetypal text within certain limits (as detailed in 2.2), but it also has other uses, such as displaying known information about the process of transmission of a text in a compact and formal way. In all of this, it is important not to forget that a stemma is always only a hypothesis, a map that must not be confounded with the mapped territory, as the above definition stresses.

Intermediate witnesses, that is, those that are lost and had exactly one descendant that is extant or gave rise to extant witnesses, cannot be depicted in a stemma unless their existence can be proven (which is rarely the case). Behind every line in a stemma, therefore, many lost intermediary witnesses may be hiding. Usually, one stemma is attempted for an entire text, but there can be cases where different stemmata for several parts of a text are necessary. In strongly contaminated transmissions, variant stemmata, that is, one stemma per *locus criticus* (see 3.3), are sometimes drawn (see 4.2.3.6). Any oriented tree has exactly one root (see 5.2), and the rest of the tradition represented by the tree descends from this root. This root is called the archetype in stemmatology. A stemma that is not a tree but only tree-like (and oriented) may have more than one root. In the case where parts of texts coalesced from various sources, the stemmata of the various components may grow into a "forest" of stemmata attached to one another with several roots. Moreover, indirect witnesses (discussed in 3.2) may provide evidence of texts that were the archetypes earlier in the textual history, before further loss of witnesses, but whose existence can now only be glimpsed in certain passages that happen to be transmitted indirectly. How to deal with such cases in a critical edition can be a difficult question, especially if the text changed significantly between earlier text-states only known incompletely from indirect witnesses and the archetype of the text as it now exists. This situation can arise, for example, in practical, fluid texts such as Latin medical texts of the Middle Ages. In general, it may be better to avoid "patchwork" editions if the text is of a rather fluid nature, and just to indicate the available older readings in an apparatus.

The stemma must not be confounded with what has, since Fourquet (1946, 5), been called the "real tree" or "complete tree" ("un arbre généalogique réel, complet"), by which is meant the (hypothetical) true genealogical tree of all witnesses of a text that have ever existed, including the lost ones (see Trovato 2017, 44–46). This entity is, of course, purely theoretical, as it would contain information that is no longer available (e.g. witnesses that are lost without a trace). In contrast, in a

stemma, only known or traceable witnesses ("junctions" in the tree) can figure. These are often only a small minority, "rari nantes in gurgite vasto" [rare shipwrecks afloat in a raging surge], as Guidi and Trovato (2004, 11) nicely quote from Virgil (*Aeneis* 1.118). The designation "real tree" would seem doubly unfortunate, for this "real" object is completely hypothetical and, moreover, does not have to be a mathematical tree at all, as there may be cycles depicting contamination. This is a typical case of technical terminology from two fields being incompatible. The computer simulations by Weitzman (1982) show nicely how the position of the archetype shifts in the "real tree" as branches are made to die off probabilistically (see fig. 2.4-2 above). This concept will be studied from a modelling point of view below (4.2.3.4).

Examples

A few examples of increasing complexity will illustrate the kinds of stemmata one can expect to encounter in editions of texts. Figure 4.1-2 shows an old stemma (1881) that is completely binary. The archetype is called "Fort.", indicating the author's name and thus failing to differentiate between the original and the archetype. Figure 4.1-3 is another old stemma (1917) with two archetypes, or to be more precise, an original and a text reworked by the author that each led to further copies. Extant witnesses are shown by capital letters and numbers. Lost intermediaries are represented by lower-case letters, where today Greek lower-case letters would be more typical (at least in classical philology). This editor chose the manuscript sigla in a

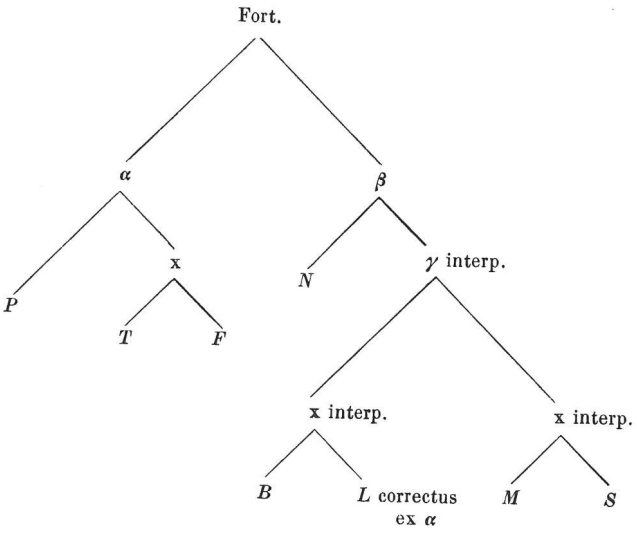

Fig. 4.1-2: Stemma for Venantius Fortunatus, *Opera poetica*, edited by Leo (1881, xxiii). Some lost intermediaries were "interpolated", that is, contaminated (see 4.4 below); manuscript *L* was corrected from a manuscript from family *α* and is thus also contaminated. Today, this would more usually be shown by a dotted line between *α* and *L*.

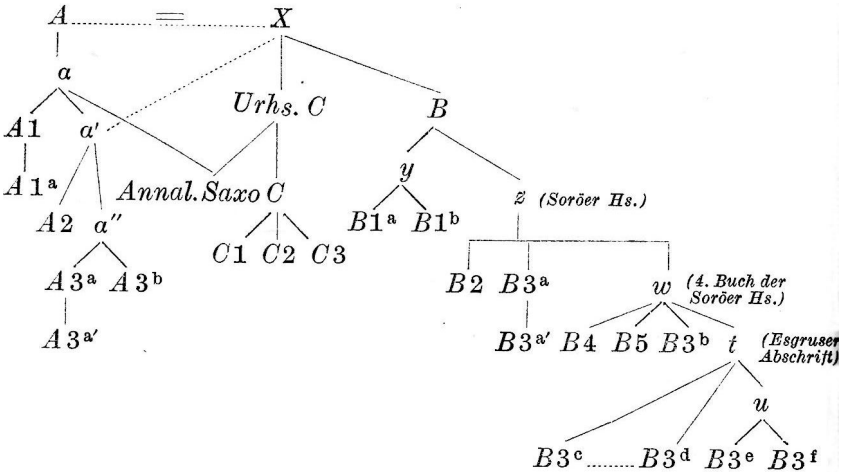

Fig. 4.1-3: Stemma of Adam of Bremen's *Gesta Hamaburgensis ecclesiae pontificum* by Schmeidler (1917, xxxiv). The author reworked his text (from *A* to *X*); *α* represents the archetype of the descendants of the first recension, *A*. Apparently, another author (Annalista Saxo) used both *α* and "Urhs. C" as sources for his own work.

way that fits their stemmatic relations (group *A*, group *B*, group *C*) in his stemma. The problem with this approach is that the sigla will have to be changed if his groups are proved wrong, thus causing confusion. Today, more neutral sigla, often indicating the present location of manuscripts – such as *V1*, *V2* for Vatican ones, or *P1*, *P2* for ones in Paris – are usual. The dotted line leading to *a'* indicates contamination, a convention that is still usual today.

A complicated modern stemma (2011) is depicted in figure 4.1-4. Lost intermediate witnesses are shown in lower-case Greek letters, extant manuscripts in upper-case Latin ones. Dashed lines represent contamination, except for the one between Ω^1 and Ω^2: Ω^1 represents the archetype, which was corrected after having been copied and gave rise to a Carolingian vulgate text (see 4.1.6), here named Ω^2. As the text became widely read in Carolingian times, this corrected, more intelligible vulgate text influenced nearly all extant manuscripts. Those older than Ω^2 were corrected ("pc" stands for *post correctionem*). Exceptions are only *A* and *W*. The estimated age of the witnesses is provided on the left-hand side.

Finally, figure 4.1-5 shows a stemma of a manuscript tradition with an extant autograph of the author (*R*; Reims, Bibliothèque municipale, 875). The original text was enlarged several times, producing extant manuscripts *B* and *P*. Sheldon-Williams believed both enlargements to be by the author and consequently based his edition on the most recent one, *P* ("Periphyseon C"); the most recent editor, Jeauneau (1996–2003), disagrees. Since the question is complex, Jeauneau decided to edit the three different versions in parallel, including his own edition in a fourth column.

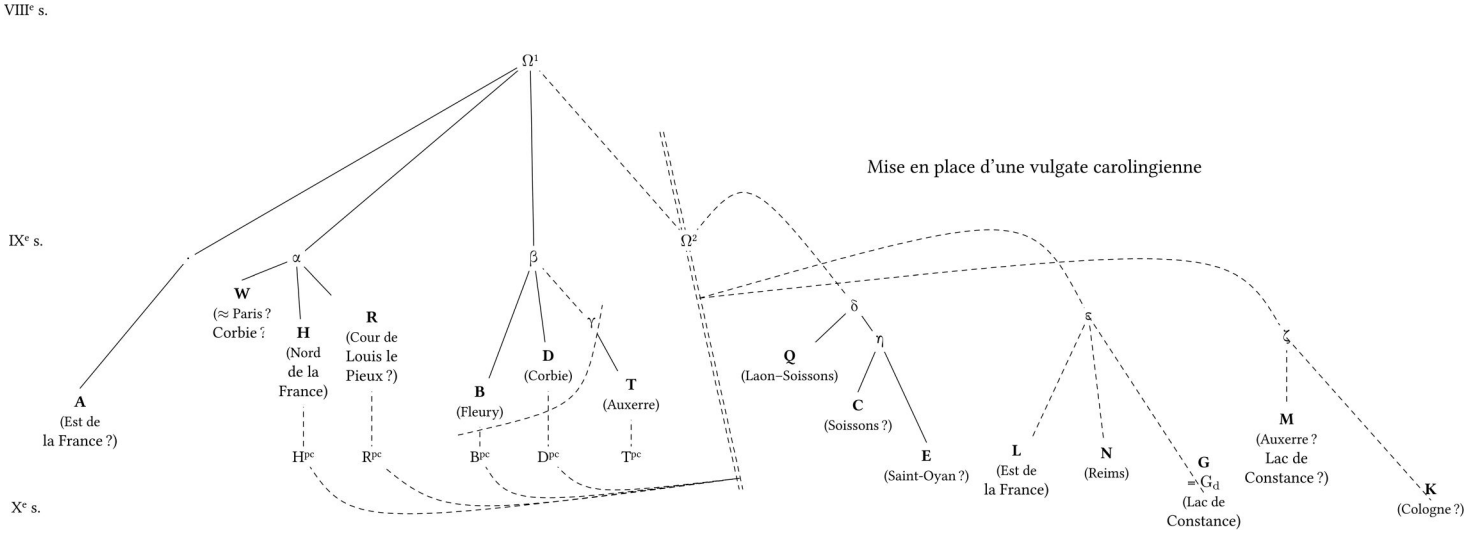

Fig. 4.1-4: Example of a complicated, modern stemma: Martianus Capella, *De nuptiis Philologiae et Mercurii*, proposed by Jean-Baptiste Guillaumin (Guillaumin 2011, cxv; slightly reworked by Guillaumin for this book).

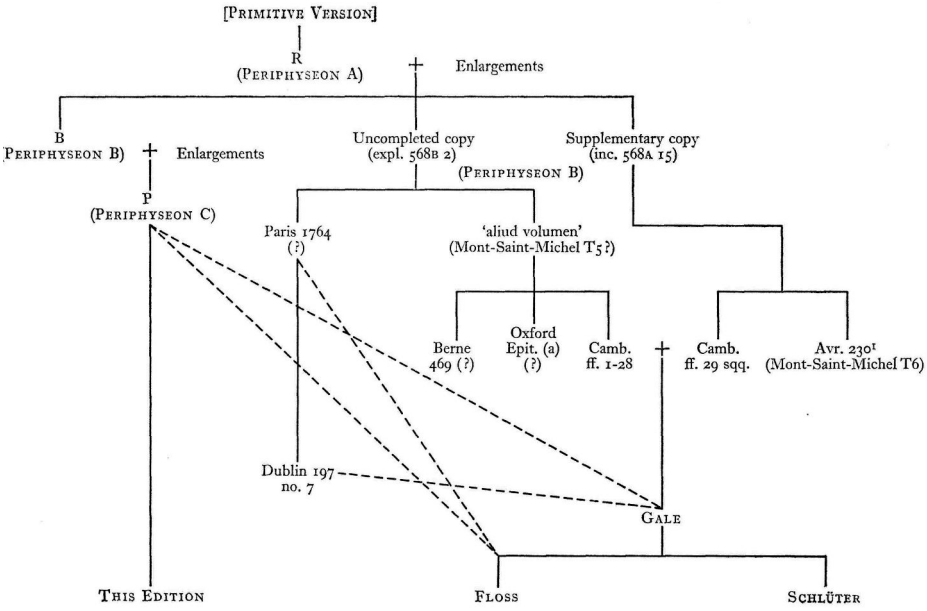

Fig. 4.1-5: Stemma of John Scottus Eriugena's *Periphyseon* by Sheldon-Williams (1968–, 1:29). The text was twice enlarged. Again, contamination is shown by dashed lines; printed editions are named in small capitals. The first print (by Gale) used a composite manuscript whose text source changed at folio 29. © Dublin Institute for Advanced Studies (DIAS).

4.1.4 Branching in stemmata

We have already mentioned above some of the possible types of stemmata one may encounter: there may or may not be contamination, and they may display one or more text-strata (as figs 4.1-3–5 above do). Others, such as variant stemmata (4.2.3.6) or the mathematical concept of the Greg tree (4.2.3.3), will be encountered in the next section. Here, however, we will address one classification scheme that has been the focus of debate for quite some time: the question of bifurcation in stemmata. In the wake of Bédier (e.g. 1928, 11), one often speaks of bifid, binary, bifurcating, or bipartite stemmata. All these adjectives derive from Latin, contain the prefix *bi-*, "two", and have related meanings. "Bifid" is derived from Latin *bifidus*, "divided into two parts"; *bipartitus* is a Latin synonym for *bifidus*; and *binarius* means anything "that contains or consists of two". "To bifurcate" stems from Latin *bifurcus*, "having two prongs or points" (all Latin meanings from Lewis and Short 1879).

A bifid stemma is a stemma in which the archetype produces exactly two branches, out of which the entire extant transmission derives. The term was first used by Bédier, who observed that, in the field of Old French manuscript traditions, almost all stemmata he encountered were bifid; this led him to question the validity

of the Lachmannian approach (cf. Bédier 1913, 1928; see 2.3–4 above, 7.3 below). Bédier speaks of a "silva portentosa" [monstrous forest] of nearly exclusively bifid trees he had found. Several theories have been proposed to explain or rationalise this phenomenon (starting with Bédier himself); they are based partly on alleged forms of mediaeval text transmission, partly on statistics, partly on psychological grounds. In the latter case, it is argued that editors tend to continue trying to find conjunctive errors until they end up with only two families, in the process possibly mistaking some shared, but polygenetic innovations for conjunctive errors. This has the convenient side effect for the editor that he must (and therefore: may) choose between the two families' divergent readings, instead of following the criterion, which would be automatic in most cases, of choosing the reading of the majority of families. The psychological argument thus amounts to the idea that the editor wishes to have some freedom in determining his text. On the other hand, Guidi and Trovato (2004) have argued, based on computer simulations, that the higher the loss rate of witnesses, the more likely bifurcations become. They tried to estimate loss rates for some early prints of which the original number of copies is known. These tend to be very high (90–100 %). Weitzman (1987, 303) had already written, referring to his own simulations, that "the present model, for example, overturns Bédier's assertion that the majority of stemmata cannot be two-branched". Hoenen, Eger, and Gehrke (2017) put forward a mathematical argument that bifurcating stemmata are indeed the most common kind of stemmata. A further critical discussion of Bédier's points can be found in Reeve (1986).

A glance at the many (and often complicated) stemmata printed in volume 1 of the *Geschichte der Textüberlieferung* (Hunger et al. 1961–1964) seems to indicate that bifid stemmata are much less prevalent for classical (Greek and Latin) texts; this impression is confirmed when looking at some mediaeval Latin editions printed in the Corpus Christianorum continuatio mediaevalis collection. It would be interesting to examine whether these differences are due to the much more standardised classical languages, or different circumstances for the transmission of the texts, or even to different approaches by the editors. In a recent study of stemmata in Old Norse philology, Haugen (2016) arrived at figures for bifid stemmata that were very similar to those in the Old French tradition. The phenomenon needs to be studied further, especially taking into account different kinds of textual traditions (different languages, different witness survival rates, different timespans between original and archetype, and the like). For now, however, it seems safe to say that Bédier uncovered a real phenomenon and not, as he believed, an artefact of a method that does not work.

In contrast to a bifid stemma, a b i n a r y s t e m m a or tree is one composed exclusively of nodes with either two children or none (not only, as in bifid stemmata, on the top level). Although real traditions of this kind of any magnitude are unlikely, one not infrequently encounters binary stemmata in editions (e.g. figs 4.1-2, 4.1-6; many examples are also printed in Bédier 1928), and many types of software (see 5.3)

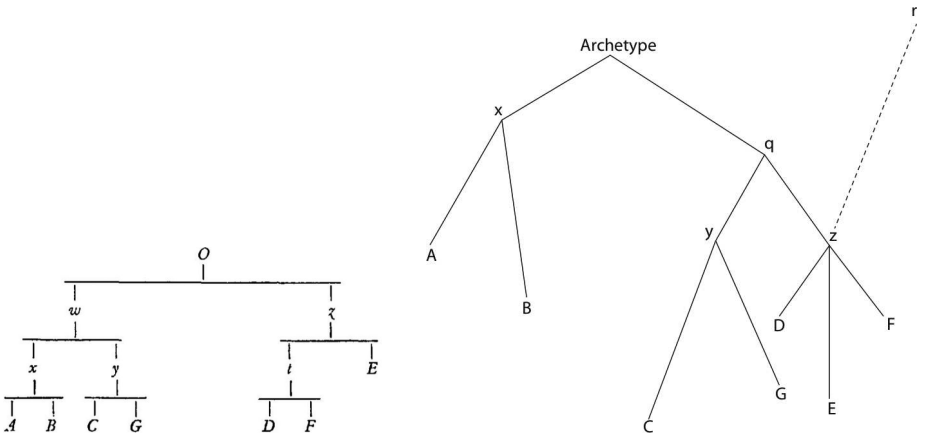

Fig. 4.1-6: (Left) the binary stemma initially proposed for the *Lai de l'ombre* by Bédier (1928; repr. 1970, 6). Bédier later accepted the criticism of Gaston Paris (1890) and modified the stemma to make it tripartite (by moving *E* directly below the archetype) before giving up on the stemmatic method. (Right) according to Trovato (2017, 294), the problem of a lacuna correctly filled in *z* can be solved by assuming extra-stemmatic contamination (see "1963" in 2.4.2). Of course, both *r* and the archetype go back to the original (not depicted).

can by design only produce binary trees, which, however, can easily be remedied by contracting nearby bifurcations into a single node (see 5.2 for more details). "Bifurcating" is a synonym for "binary" in manuscript studies, whereas "bipartite" may be used as a synonym for "bifid" or "binary". On the whole, the usage of these terms does not seem to be fully fixed yet.

There are, however, also many stemmata with a lot more than two branches issuing from the first node (the archetype). Figure 4.1-7 shows such a case: the stemma of Petrus Alfonsi's *Dialogus*, exhibiting eight branches directly from the archetype. This case, probably quite rare, of such a high initial filiation (the archetype is close in time to the original – indeed, the two may be identical in this case) is explained by the fact that the book immediately gained great popularity. It may be that the author, who was a travelling scholar, frequently left his abode, which will have made it likely that local disciples wanted to keep a copy (on this hypothesis, see Cardelle de Hartmann, Senekovic, and Ziegler forthcoming, chap. 1).

In graph theory, the term "bipartite" means something entirely different. There, a bipartite graph is a graph whose nodes can be arranged into two disjoint sets such that every edge connects a node in one of them to one in the other (i.e. in each of the two sets, there are no nodes that are connected with one another; Diestel 2005, 17). It can be proved by induction that every tree is a bipartite graph (in this mathematical sense). These two meanings of "bipartite" should not be confused.

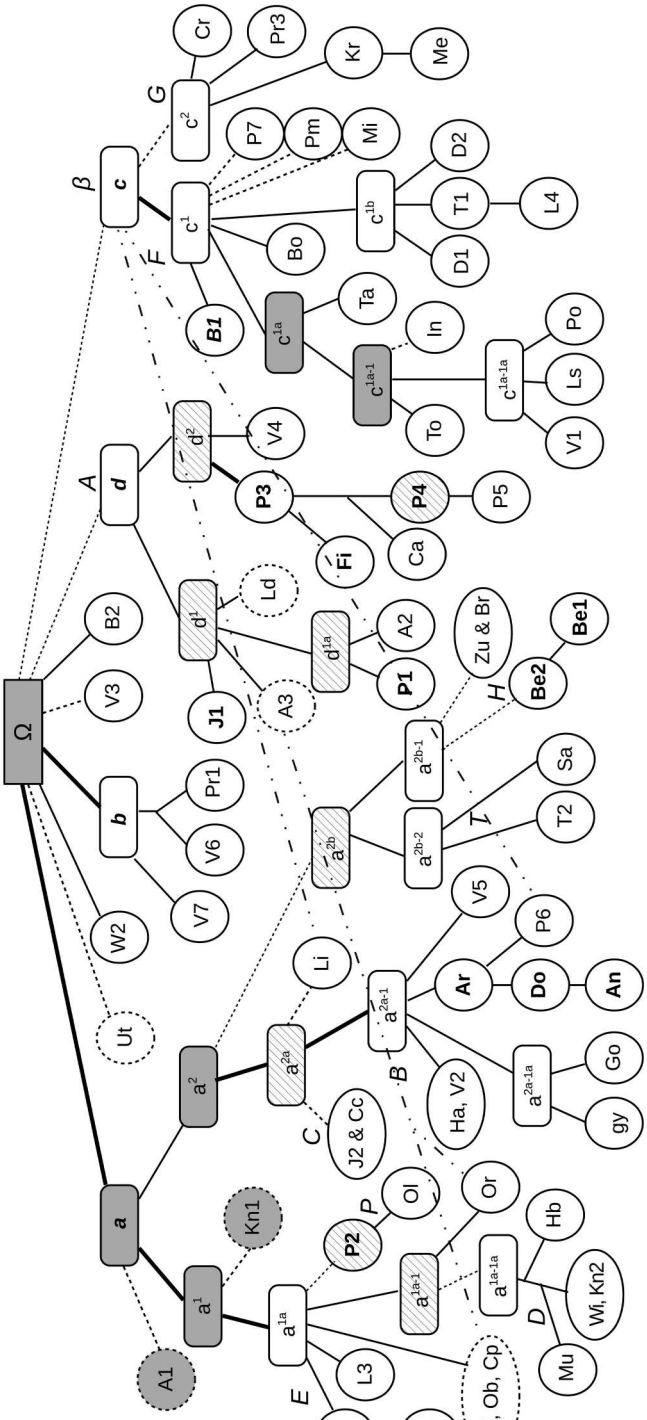

Fig. 4.1-7: A complicated stemma for Petrus Alfonsi's *Dialogus* with strong filiation from the archetype (from Roelli 2014, 55). Thick lines indicate few changes, dashed ones many. Dot-dash lines indicate contamination.

4.1.5 The archetype

Below, it will become clear that the term "archetype" is used in some slightly differing ways today. We would propose the following definition: the archetype is the most recent witness from which all extant witnesses of a text derive.

It follows from this definition that the archetype's text is as close to the original state of the text as the surviving witnesses can attest. According to this definition, the archetype may in some cases be identical with the original – if the original itself has survived, or if more than one copy of the original has produced extant offspring. For classical or early mediaeval texts, however, this is very rare. An example of a text from the ninth century that has come down to us in the original is the *Periphyseon* by John Scottus Eriugena (Jeauneau and Dutton 1996; see also 1.2.4, and fig. 4.1-5 above). At any rate, the concept of an original is stronger than that of an archetype; in other words, if an archetype of a text can be shown to have been the original, it is usually addressed as the "original" and treated accordingly. For texts from Antiquity or the Middle Ages, the low chances of having an extant archetype are still somewhat higher than those of having an extant original. If the archetype is not extant, one of the aims of *recensio* (see 6.2) is to reconstruct its text as far as possible. Insofar as it has become the archetype by means of historical accident, this witness may have borne a corrupt text and may have been written by an incompetent scribe; in order to arrive at a readable text, the editor may have to resort to *emendatio* (see 6.2.2.1). On the other hand, it may happen that an especially authoritative copy becomes the archetype because other less authoritative copies are discarded or not copied further (see 4.1.6). The quality of the archetype may be an important parameter for gauging the kind of *arbre réel* one has to expect for a textual tradition. For instance, for Varro's *De lingua Latina* we have an extant but very corrupt archetype from the eleventh century containing five of the original twenty-five books (Firenze, Biblioteca Medicea Laurenziana, li.10).

In a stemma, the archetype is placed immediately below the original (if the latter is depicted at all) and, especially in classical philology, it is often denoted by a Greek letter. Figure 4.1-8 shows the path between the original (*X*) and the archetype (*α*), which may have consisted of many and complex branches, all of which are completely lost, as a mere line. As we have seen above in figure 4.1-3, for some works, more than one version of the original may have to be reckoned with (e.g. if the text was reworked by the author); for similar reasons, more than one (state of the) archetype may exist (as in fig. 4.1-4). Several states of an archetype can arise if it was reworked and marginal or *supra lineam* variants were added to it. This will make the reconstruction of the stemma more difficult, as some copyists may simply have omitted these variants while others may have incorporated them (or some of them) into the text while omitting the original readings. In certain traditions (especially very contaminated or fragmentary ones), it may be impossible to arrive at an archetype.

A hyparchetype is a state of the text, often but not necessarily lost, which is situated directly below the archetype in the stemma. The word is derived from

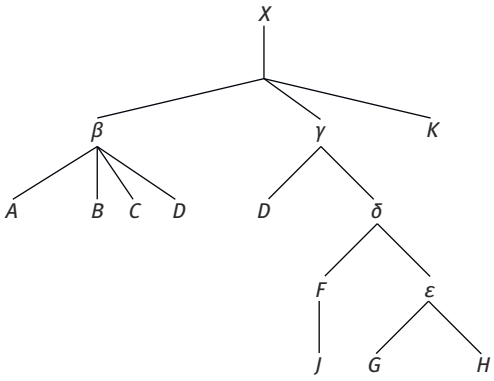

Fig. 4.1-8: Model stemma from Maas (1960, 7), redrawn and slightly simplified.

Greek ὑπό, "under, below", and ἀρχέτυπον (see above). Thus, in figure 4.1-8, *β* and *γ* are lost hyparchetypes, and *K* is an extant one. The term is also occasionally used more loosely for ancestors of families that do not go back directly to the archetype, such as *δ* in our example. Hyparchetypes are thus the ancestors of related families of preserved witnesses. Like the archetype, hyparchetypes are often denoted by Greek characters in the stemma, especially in classical philology. Paul Maas proposed using the term "hyparchetype" in a more exclusive sense to refer to reconstructed variant-carriers (1960, 8), that is, lost witnesses directly below the archetype. He considers as variant only those readings directly below the archetype between which no mechanical choice is possible. The alternative form of the term, "subarchetype" – with *sub*, the Latin synonym for ὑπό – is not recommended, but is sometimes found in the literature.

There are many subtly different definitions of the key concept of the archetype in the literature. Reeve (1986) collected a list of about a dozen such definitions, some (but not all) of which are identical or equivalent to the above definition. In particular, there is contention about two points. First, it is disputed whether an extant archetype should still be called an archetype. It may be argued that in such a case all other witnesses can be eliminated (*eliminatio codicum descriptorum*; see 2.2.8) and – at least for the reconstruction of the primordial text – the situation becomes equivalent to that of a *codex unicus* (see 3.3.2) without an archetype. For instance, Montanari would in such a case speak of a "codex unicus secondario" (2003, 21). Other opinions differ; Pasquali, for example, was content to address an "archetipo conservato delle Metamorfosi di Apuleio" (1934, 33) [the extant archetype of the *Metamorphoses* by Apuleius], and we would prefer to speak at least of a trivial archetype in such cases. Second, in the cases where the same witness is both original and archetype, it will in most cases likewise make little sense to speak of the original as an archetype, and some authors would altogether avoid this. From a practical philological point of view, it is indeed preferable to avoid doing so, as the text has to be treated differently depending on whether it is sanctioned by the author (in the case of an original) or a product of historical chance

(non-original archetype); but from a graph-theoretical point of view, both are MRCAs (most recent common ancestors) and stand in need of a common designation. The existence of an archetype different from the original can be proved by finding at least one error common to the entire tradition, one the author could not have written. One may, therefore, differentiate between a narrower concept of the archetype which excludes originals and extant archetypes and which is especially useful in the context of ecdotics, and a wider, purely positional one that equates "archetype" with "MRCA". Here, we follow the latter.

In graph-theoretical language, finding the archetype is equivalent to assigning a root in an unrooted tree (see further 4.2). In evolutionary biology, the term "most recent common ancestor" (MRCA) is used similarly to "archetype" in textual criticism. Here, however, the similarity to phylogenetics ends: the concept of an original makes little sense in biology, unless one chooses to go all the way back to the so-called LUCA (last universal common ancestor) of all living beings, to which, however, no urtext of all existing texts ever written can be compared.

No matter whether computerised or traditional approaches are used, deciding where in the tree the archetype is to be located is often the most difficult, but also the most crucial, task for a philologist studying a textual tradition with an interest in the original text. In the traditional method, the problem is usually less pronounced because good significant errors (see 2.2.5) can often be identified. They are, in Greg's terminology, "substantive variants" (also known as "substantial" ones), which he defined thus:

> we need to draw a distinction between the significant, or as I shall call them 'substantive', readings of the text, those namely that affect the author's meaning or the essence of his expression, and others, such in general as spelling, punctuation, word-division, and the like, affecting mainly its formal presentation, which may be regarded as the accidents, or as I shall call them 'accidentals', of the text. (Greg 1950–1951, 21)

A subclass of such substantive variants – those that cannot be undone by an intelligent scribe – can serve as significant errors. These tend to be directed; that is, the editor can determine which variant is original (or at least archetypal) and which one(s) are innovated. There are several aids at the philologist's disposal for this task: old ones such as *lectio difficilior* (see 4.3.2), as well as more recent ones such as diffraction (see "1955–" in 2.4.3). In order to do this correctly, knowledge about the text and its author, or the archetype and its scribe, must be inferred and used. Computerised approaches from biology are not usually helpful for this, as biologists tend to use an outgroup to root their trees. The outgroup, as will be explained more fully below (5.2.1), is an organism distantly related to the group of taxa being studied. The point where its branch exits the tree then corresponds to the MRCA of the studied group. As texts are written at some point in time *ex nihilo*, so to speak, this approach cannot usually be used for rooting the tree (see 3.2.8, 4.5.3 for exceptions). The example in figure 4.1-6 above shows how a stemma can change radically if another node in the tree is designated as the archetype. Such changes lead to a very

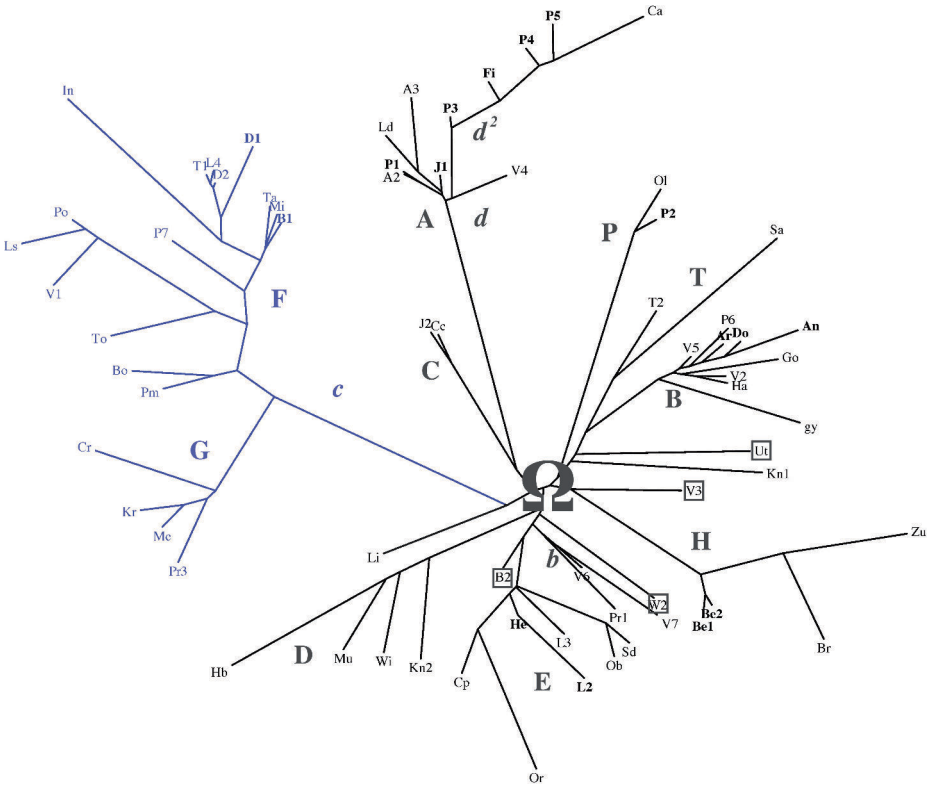

Fig. 4.1-9: How does one find the archetype in an unoriented tree? Example from Roelli (2014, 47) with added blue colour. The tree was drawn fully automatically; only the superimposed letters naming the families were added manually. The boxed witnesses are the oldest ones.

different influence of the witnesses on the reconstruction of the archetypal text. In Bédier's stemma, *E* is the most important witness (with a weighting of 25 %), whereas in Trovato's, *A* and *B* are (with a weighting of 25 % each). Once the philologist, perhaps using software, has arrived at an unrooted tree-graph of the relationships between all extant witnesses, direction in the tree must be provided by discerning for some variants which one is original or archetypal and which one(s) is (or are) innovated. A priori, the archetype may be hiding at any point in the tree, even perhaps on an edge between two nodes. An example from the recent edition of Petrus Alfonsi's *Dialogus* will illustrate the approach. "*Leitfehler*"-based software (as described in 5.3.7) produced the unrooted tree depicted in figure 4.1-9. In the following passage, the text marked "°…°" is missing in all witnesses of the *c* group (blue in fig. 4.1-9):

> Nunc cognoscere potes quia gradus signi qui est in oriente sole Aren ciuitati apparente °non est idem cum eo qui eadem hora alii ciuitati apparet. Similiter gradus qui est in occidente sole in Aren occumbente° non est idem cum eo qui eadem hora alii apparet ciuitati. (Cardelle de Hartmann, Senekovic, and Ziegler 2019, § 56)

Philological judgement is required to observe that the omission is best explained through eye-skip ("non est idem [...] non est idem"; see 4.3.2 for more on this phenomenon) and that, therefore, the *c* group has innovated by accidentally removing the words in question. It follows that the blue parts of the tree cannot contain the archetype. Similar arguments show that, in this case, the archetype is indeed in the middle of the plot (marked as Ω). If editors fail to conduct this step of determining the direction for some variants (the "significant errors" discussed in 2.2.5), it is likely that a wrong place in the tree will be chosen as the would-be archetype. This may be the most "neutral" text or the one most commonly read in some key period, possibly long after the archetype. This leads into the consideration of such textual vulgates.

4.1.6 Vulgates

To conclude this section, we look at a concept related to that of the archetype, that of a textual v u l g a t e. The word derives from the Latin *vulgata*, "spread among the multitude (*vulgus*)"; the feminine noun *editio*, "edition", is implied, so the form is feminine. The same term is more frequently met in biblical studies, but there it denotes something completely unrelated: St Jerome's Latin translation of the Bible, which became the most widely used one in the Middle Ages and beyond. In textual criticism, a vulgate text is the text form that reached the widest distribution at a time, possibly long after the archetype, when interest in the text experienced an upsurge for one reason or another and many copies were made. When interest in a text is high, it is also likely that some people will compare witnesses in order to arrive at a better text. This is, in fact, nearly the same thing that modern philologists using the genealogical method do, although before the nineteenth century the scientific tools for arriving as close to the archetypal text as possible were not yet in existence and the result depended a lot more on the editor's intuition. Vulgate texts are thus often a kind of early text edition or, to put it negatively, the product of heavy contamination. Their text may supplant all other text forms and thus eradicate them. Trovato (2017, 299–333) provides an example in his discussion of the transmission of Dante's *Divina Commedia*. A v u l g a t e r e a d i n g is a reading present in a vulgate; it can also refer simply to the most frequent reading, and often implies that this reading is not the original one.

If witnesses are grouped based on all undirected variants, instead of exclusively on directed common errors, as might be the case on the part of inexperienced textual editors using software methods, there is a great risk of arriving at a vulgate text instead of the archetype (see Trovato 2017, 138–144). In some cases, it will make sense to edit a vulgate text because it was the most frequently read one, but it is important to be aware of the difference between vulgate and archetypal texts.